An Experimental Study of Cohesion and Coupling Metrics

Chung-Horng Lung
Department of Systems & Computer Eng.
Carleton University, Ottawa, Ontario Canada
email: chlung@sce.carleton.ca

Nishith Goel and Pulak Dhar
Cistel Technology Inc.
Ottawa, Ontario Canada
email: ngoel@cistel.com and pulak@cistel.com

Qualities Related to Cohesion and Coupling

- Survivability
- Maintainability
- Verifiability
- Flexibility
- Portability
- Reusability
- Interoperability
- Expandability
Metrics & System Studied

Metrics:
• H. Dhama’s C&C metrics with modifications
  *J. of Systems and Software* 1995
• D. Card and R. Glass’s Data Complexity
• Relative Complexity
• LOC (Lines of Code)
• Cyclomatic Complexity
• LSS (Logical Source Statement)
• Others: Average Depth, Nesting Level, etc.

System:
  Large-scale telecommunications software system

Dhama’s Cohesion & Coupling Metrics

• Functional Cohesion
• Data Flow Cohesion
• Action-Bundling Cohesion
• Logical Bundling Cohesion
• Coupling
C&C Metrics Used in the Study

- Functional Cohesion: slight modification
- Data Cohesion: hybrid of data flow & action-bundling cohesion
- Uses Cohesion: density of variables
- Coupling: slight modification
- Data Complexity
- Logical Bundling Cohesion: similar to logical nesting level, not used.

Functional Cohesion

\[ F = \frac{1}{p}, \]

where \( p = i_1 + q_1 i_2 + u_1 + q_2 u_2 + l_1 + q_3 l_2 + g_1 + q_4 g_2 + q_5 w \)

- \( i_1 \): in data parameters
- \( i_2 \): in control parameters
- \( u_1 \): out data parameters
- \( u_2 \): out control parameters
- \( l_1 \): number of local variables used as data
- \( l_2 \): number of local variables used as control
- \( g_1 \): number of global variables used as data
- \( g_2 \): number of global variables used as control
- \( w \): number of modules called

\( q_1, q_2, q_3, q_4, \) and \( q_5 = 2 \)

A variable could be used as both data & control.
A parameter could be both in & out parameter.
Data Cohesion

- Interdependencies among the different statements depending on the processing of data.
- Hybrid of Dhamá’s Data-flow and Action-bundling cohesion.
- Consider the type of statement and position of the variable.
- Data-flow cohesion occurs between two statements if data used in one statement is transformed and then used by another transformation or data in another statement.
- Action-bundling cohesion occurs when several actions are performed on a single piece of data which results in that data being transformed.
- Data cohesion (hybrid of data-flow & action-bundling) measures the number of statements (distance) separating pairs of statements that have either cohesion.

Data Cohesion

Example:

Data flow cohesion: B is transformed and used in another statement.

\[
\begin{align*}
B &= A \\
\ldots &= \ldots \\
C &= B \\
\text{if } (B < D) &\quad \text{write } (B)
\end{align*}
\]

Action-bundling cohesion: C is used in both statements on the right.

\[
\begin{align*}
B &= C \\
\ldots &= \ldots \\
D &= C \\
\text{for } i \text{ in } 1..C &\quad \text{write } (C)
\end{align*}
\]
Uses Cohesion

- Uses cohesion measures code density and involves the number of local variables and global variables divided by the number of tokens (variables + constants + function calls) in the code.

Coupling

\[ F = \frac{1}{p}, \]

where \( p = i_1 + q_1i_2 + u_1 + q_2u_2 + g_1 + q_3g_2 + w + r \)

- \( i_1 \) = in data parameters
- \( i_2 \) = in control parameters
- \( u_1 \) = out data parameters
- \( u_2 \) = out control parameters
- \( g_1 \) = number of global variables used as data
- \( g_2 \) = number of global variables used as control
- \( w \) = number of modules called
- \( r \) = number of modules calling the module under consideration

\( q_1, q_2, q_3, \) and \( q_4 = 2 \)

A variable could be used as both data & control.
A parameter could be both in & out parameter.
Data Complexity

- Data Complexity

\[ D(i) = \frac{V(i)}{f(i)} + 1 \]

where

- \( D(i) \) = data complexity of module \( i \)
- \( V(i) \) = I/O variables in module \( i \)
- \( f(i) \) = fanout of module \( i \)

Each point represents a module which consists of one of more files.
Conclusions

- Functional cohesion, data complexity, and coupling seem to be consistent enough to produce detectable trends.
- LOC and Cyclomatic complexity seem correlate well with functional cohesion and data complexity.
- Other metrics studied that also correlate well with functional cohesion and data complexity include
  - Logical Source Statement
  - Physical Source Statement
  - Nesting Level
  - Average Depth